DOCUMENT RESUME

ED 220 092 IR 010 312

AUTHOR Khosh-Khui, Abolghasem
TITLE Statistical Analysis of the Association Between Subject Headings and Their Corresponding Class Notations in Science and Technology Monographs.
PUB DATE [81]
NOTE 40p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Classification; Correlation; Indexing; Library Research; Regression (Statistics); *Subject Index Terms; Tables (Data)
IDENTIFIERS Dewey Decimal Classification; Library of Congress Classification; *Scientific and Technical Information

ABSTRACT

This study investigates the degree of relationship between scientific and technical subject headings and their corresponding class notations in the Dewey Decimal (DDC) and Library of Congress Classification (LCC) systems. The degree of association between a subject heading and its corresponding class of notation or notations is measured by counting the frequency of occurrence of that subject heading and comparing this score with the most frequent identical class notation for the same subject heading. The correlation between these scores, i.e., the probability of co-occurrence of a subject heading and a class notation, and the number of subject headings are the two measures of association used in the study. The study considered 3,280 subject headings and class notations in 1,893 bibliographic records. The results show that as the number of subject headings per bibliographic record increases, the degree of association between subject headings and their corresponding class notation decreases. Except for subject headings with the order of one, LC and DDC were not significantly different. There are 20 references. (Author/RAA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED220092

STATISTICAL ANALYSIS OF THE ASSOCIATION BETWEEN SUBJECT HEADINGS AND THEIR CORRESPONDING CLASS NOTATIONS IN SCIENCE AND TECHNOLOGY MONOGRAPHS

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ABOLGHASEM KHOSH-KHUI*

School of Library and Information Science, Indiana University
Bloomington, Indiana.

---

* Mr. A. Ghosh-khui is a doctoral candidate for Library and Information Science, at Indiana University.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. Khosh-Khui

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR010312

STATISTICAL ANALYSIS OF THE ASSOCIATION BETWEEN SUBJECT HEADINGS AND THEIR CORRESPONDING CLASS NOTATIONS IN SCIENCE AND TECHNOLOGY MONOGRAPHS

Abstract--3,280 Subject headings and class notations in 1,893 LC bibliographic records were studied to determine changes in the degree of the association between subject headings and their corresponding class notations, with respect to the changes in the number of subject headings per bibliographiic record, and the order of subject headings in those records. The results showed that as the number of subject headings per bibliographic record increases, or as order of subject headings increases, the degree of association between subject headings and their corresponding class notations decreases. Except for subject headings with the order of one, the Library of Congress and Dewey Decimal Classifications were not significantly different with respect to the measures of association in this study.

INTRODUCTION

Classification and subject headings are two different approaches to subject matter in a collection. Shelf classification provides a systematic approach to subject matter of documents, while subject headings provide an alphabetical approach. Classification schemes and subject headings together provide library patrons with a dual approach to library materials, the classified and alphabetic.[1] According to Daily, ideally class notations should be used for the purpose of shelf arrangement as well as symbolizing subject matter.[2] He notes the different purposes of these two systems of organizing knowledge: "Headings are meant to bring together what the classification separates."[3] He also argues that the 'ideal scheme of the

arrangement of knowledge' is the object of subject heading while 'utilitarian arrangement of books on a shelf' is the purpose of classification.

Manheimer states that, " the function of the classification scheme is to provide a place for every work in the collection, in which that work will be shelved with other works that are alike according to the scheme." [4] She points out the fact that each book can receive only one class number because it can be shelved in only one place, but theoretically any book can receive as many subject headings as there are subjects in the book; though in practice, assignment of subject headings is usually limited.[5]

It should be noted that not all libraries use classification only for the purpose of shelf arrangement. Some libraries and institutions in the United States do use classification for purposes other than what Fairthorne[6] called the 'marking and parking' syndrome. The Universal Decimal Classification is used, for example, at the National Center for Atmospheric Research in Boulder, Colorado, and its professionals in that institution manage to shelve their books while assigning more than one classification per item.[7]

Therefore, while in a particular classification system there may be more than one subject heading in a bibliographic record, there is usually only one class notation corresponding to all of

the subject headings in that record. The policy statement of the Library of Congress[8] is precisely directing that regardless of the number of subject headings, the first subject heading should represent the class notation:

> "the first subject heading assigned should normally represent the actual class number assigned to the work(the monographic class number for analytics of collected sets). In some cases a combination of two headings is needed to represent the class number, and these two should normally be listed as the first and the second subject headings assigned. Important headings which narrowly miss designating the class number (such as a special approach to the major topic, or the second of two major topics of the work) should be listed next. Finally, headings added to elaborate certain aspects, such as bibliographical headings, etc., listed last...".

This statement of classification policy implies that for any bibiliographic record, the degree of association between the class notation and the first subject heading will be greater than the the degree of association between it and the second, third, or fourth subject headings. The statement also suggests that the degree of association between the class notations and subject headings in 'single subject heading' bibliographic records will be greater than the degree of association between class notations and subject headings in 'multiple subject heading' bibliographic records.

Given a classification system and a subject heading list, what would be the chances that all the bibliographic records with a given subject heading such as 'Electric Welding' will have the

same class notation, '671.521' in Dewey Decimal Classification [9], or 'TK4660' in Library of Congress Classification[10]? If these subject headings do not have the same class notation, how can such variation be explained? Is the variation explained by the bibliographic records which have more than one subject heading? If so, does the order of headings have an effect? Alternatively, could the variation be the result of changes in class notation in different editions of the classification system? Could the variation be the result of availability of more than one class notation for a given subject heading (In the present example, for subject heading 'Electric Welding', there are two corresponding class notations in DDC Schedule: '671.521' and '672.521', depending on the content of the document being classified).[11] Another possibility is cataloger and/or classifier error, or inconsistency over time in assigning class notations. Also, is variation different if a more specific subject heading is searched?

This study attempts to find the answer to some of the questions raised above by investigating the degree of relationship between subject headings and their corresponding class notations in the Dewey Decimal Classification system and between subject headings and their corresponding class notations in the Library of Congress Classification system. The degree of association between a subject heading and its corresponding class notation or notations is measured by counting the frequency of

occurrence of that subject heading and comparing this score with the frequency of the corresponding most frequent identical class notation for the same subject heading. The correlation between these scores, the probability of co-occurrence of a subject heading and a class notation, in various bibliographic records with different numbers of subject headings, are two measures of association used in this study.

REVIEW OF LITERATURE

Richmond studied the syndetic structure of 'see also' references in the Library of Congress Subject Headings explicating a concealed classification in the references involved. In her study, the organization in the subject heading's cross-references was compared with subject norm, and on the assumption that the cross reference structure should be parallel to the subject's internal classification, a logical example was demonstrated.[12] Using the subject heading 'Cats', she showed that in one type of organization, the 'see also' cross reference, there is a concealed classification inherent in it. She suggested that "the adoption of classification principles in making 'see also' references should result in a more rational and therefore more functional structure for this part of the subject catalog."

Harry Dewey[13] explored the question of the degree to which a

subject catalog could satisfactorily substitute for the three systems of classification: Dewey Decimal Classification, Library of Congress Classification, and Universal Decimal Classification. Since classification numbers are a kind of heading, he argued, some degree of substitution is possible. He attempted to show "that a numerical subject catalog and a corresponding shelf arrangement are almost, by definition, interchangeable." He suggested that both a subject catalog and a shelf classification are needed to give best services to the users.

Manheimer[14] studied the relationship of classified Library of Congress Subject Headings to the Library of Congress Classification schemes. She compared class GR (Folklore) with the Library of Congress Subject Heading List and noted that "there is a very close relationship between the subject headings and classification schemes in that almost every class heading has at least one matching or equivalent subject heading." She also found that "approximately 90 per cent of the time the subject heading appears as the topic word, and approximately 70 per cent of the time it appears unglossed with only a class reference to Folklore." However, her conclusion was somewhat contradictory with the above findings. She concluded that "in spite of the fact that there are subject headings for almsost every class heading and that there is a close relationship between the two lists, these are two different systems of subject analysis, only tangentially related and then in an entirely unpredictable

way."[15]

Immroth[16] investigated the relationship between the classification schedules and their indexing on one hand and subject heading and classification on the other hand. He found that about half of the class notations in the Library of Congress Classification schedules are indexed. He also noted that there is a significant relationship between the terminology of the Library of Congress subject headings and the Library of Congress classification.

To obtain data which reflects the current cataloging practice, O'Neill and Aluri[17] studied characteristics of subject headings occurring in cataloging records. Their study was based on 33,455 cataloging records. There were 50,213 subject headings in their study, of which 47,036, or 93.7 per cent, were Library of Congress Subject Headings. They found that topical subject headings were the most common type of subject headings in the cataloging records. Seventy per cent of subject headings were topical subjects. Sixty-two per cent of all records contained at least one topical subject heading. Geographic subject headings composed fifteen per cent of the total subject headings.

STATEMENT OF THE PROBLEM

Given a classification system and a list of subject

headings, while there may be one or more subject headings in a bibliographic record, there is usually one class notation in that bibliographic record. With regard to this difference, it is expected that, whenever the number of subject headings per document increases, there would be some variations in the degree of association between subject headings and their corresponding class notations.

(a) Research Questions

This study seeks to answer the following questions:

1. Does the degree of association between the subject headings and class notations vary, when the number of subject headins per bibliographic record increases? The above question can be precisely stated in the following two questions:

1a. Is there any change in the degree of association between the Library of Congress Subject Headings (LCSH) and the Library of Congress Classification (LCC) notations, when the number of subject headings per bibliographic record increases?

1b. Is there any change in the degree of association between LCSH and Dewey Decimal Classification (DDC) notations, when the number of subject headings per bibliographic record increases?

2. Does the degree of association between the subject headings and their corresponding class notations vary with

respect to the changes in the order of subject headings in the bibliographic record? Again, this question can be precisely stated in the following two questions:

2a. Is there any change in the degree of association between LCSH and LCC notations, when the order of subject headings in bibliographic records changes?

2b. Is there any change in the degree of association between LCSH and DDC notations, when the order of subject headings in bibliographic records changes?

3. Would the association between the Library of Congress Subject Headings and class notations in the Library of Congress Classification system be significantly different from association of the same subject headings with respect to the class notations in Dewey Decimal Classification system?

(b) Variables

There are four variables in this study. These variables are as follows:-

1. The Library of Congress Subject headings: Headings in bibliographic records used by the Library of Congress Cataloging Division. Wepsiec[18] identifies fifteen types of subject headings. However, in this study there has not been any distinction between various types of subject headings. Subjects are operationally defined as a word or any string of words or symbols used as a heading, which is different from any other

heading at least by one character. Therefore according to this definition the following four strings of characters would be regarded as four subject headings:

Building.
Buildings.
Buildings--Terminal Properties.
Buildings--Terminal Properties--Testing.

2. The Library of Congress Subject Headings Order: This variable is a characteristic of assigned subject heading and can be any number from one to four. If any bibliographic record has more than four subject headings, only the first four subject headings are included in the study.

3. Category of Bibliographic Records: This variable represents the number of subject headings per bibliographic record and can vary from one to four. Any bibliographic record with more than four subject headings is regarded as four.

4. Class Notation: This variable includes class notation for both LCC and DDC systems. The maximum number of characters or digits used for comparison is fifteen characters and/or digits.

(c) Sampling

Six MARC tapes were randomly selected from 1972 weekly Library of Congress MARC tapes available in the Research Center of the School of Library and Information Science, Indiana University. However, since a complete year of MARC tapes was not

available, the sample may not be regarded as a completely randomized sample. Subject headings, Dewey Decimal Classification notations, and the Library of Congress Classification notations of all bibliographic records with a Dewey Decimal Classification notation of 500-699 were retrieved from these MARC tapes.

Restrictions were imposed to make sure that all selected bibliographic records have at least one subject heading, and have both DDC and LCC notations. There were 8,396 bibliographic records on the six weekly LC MARC tapes, of which 1,893, or 22.55 per cent, met the above criteria and were selected.

(d) Distribution of Bibliographic Records

Of 1,893 retrieved bibliographic records, 902 or 47.64 per cent had only one subject heading, 688 records or 36.34 per cent had two subject headings, 233 records or 12.31 per cent had three subject headings, and only 70 records or 3.69 per cent had four subject headings or more. Table 1 shows the distribution of retrieved bibliographic records according to the number of subject headings assigned to them by the Library of Congress. The retrieved subject headings for each category of bibliographic records is obtained by multiplying the number of subject headings in each category by the number of retrieved records (for instance, 2 x 688 = 1376). Therefore the analysis in this study

is based on 3,280 subject headings. The suject headings from the first category of bibliographic records constitute 27.5 per cent of the total subject headings retrieved. The next three catagories of bibilographic records constitute 41.95 percent, 21.31 per cent and 9.23 per cent of the total retrieved subject headings respectively.

TABLE 1
DISTRIBUTION OF RETRIEVED BIBLIOGRAPHIC RECORDS ACCORDING TO THE NUMBER OF SUBJECT HEADINGS.

Number of Subject Headings	Retrieved Records Number	Per cent	Retrieved Subject Headings	Subject Headings Per cent
1	902	47.64	902	27.50
2	688	36.34	1376	41.95
3	233	12.31	699	21.31
4 or more	70	3.69	303	9.23
Total	1893	100.00	3280	100.00

(e) Arrangement of Data

Retrieved bibliographic records were rearranged in a single file, with appropriate codes to designate the order of subject headings in the bibliographic records, and the total number of subject headings in the bibliographic records, from which a given subject heading was retrieved. The new file was sorted into two different lists with different orders of sort keys. These sorted lists will be referred to as 'List #1', and 'List #2'. The order of sort keys in each list is given in Table 2.

TABLE 2
TWO DIFFERENT ORDERS OF SORT KEYS USED TO SORT DATA.

ORDER OF SORT KEYS	(1) 'List #1'	(2) 'List #2'
1st	LCSH	LCSH
2nd	Category	Category
3rd	LCSH Order	LCSH Order
4th	LCC Notation	DDC Notation

Each list was used for analyzing a different research question in this study. The purposes of creating these four lists were as follows:

1. To analyze the degree of association between LCSH versus LCC notations, the retrieved file was sorted according to the sort keys order given in column (1) to create 'List #1'.

2. To analyze the degree of association between LCSH versus DDC notations, the retrieved file was sorted according to the sort keys order given in column (2) to create 'List #2'.

It should be noted that, in above two lists, when the 'LCSH Order' sort key was placed before 'Category' sort key there were slight differences in the results. The order of sort keys selected in this study gives the same results as though each group of bibliographic records (i.e. bibliographic records with one subject heading, two subject headings, etc.) was separately analyzed.

ANALYSIS OF DATA

Several COBOL Programs were developed to count the frequency of subject headings and their corresponding class notations, and computing the probabilities in four categories of bibliographic records. In addition, Statistical Package for the Social Sciences (SPSS) was used to compute Pearson's correlation and the regression equations.

The following analyses were done to find out the answers to the research questions in this study:

1. Analysis of Subject Increase Per Bibliographic Record

With regard to the first research question, to see if the degree of association between subject headings and their corresponding class notations varies with respect to the increase in the number of subject headings per bibliographic record, the measure of probability was calculated. The probability measure in any group of bibliographic records was defined as follows:

$$P_{ci} = \frac{\sum f_{cm}}{\sum f_{s}}$$

Where p_{ci} is the probability of class notations to be the same for any given subject heading in category i.

ξf_{cm} is sum of the most frequent class notations, corresponding to any the subject headings which had an occurrence of two or more.

ξf_s is the sum of the frequencies of the subject headings which had an occurrence of two or more.

Therefore the probability measure, in any group of bibliographic records, was obtained by dividing the sum of the frequencies of the most frequent class notation, corresponding to any given subject heading which had an occurrence of two or more, by the sum of frequencies of the same subject headings in that group. The sum of subject heading frequencies, in turn, was obtained by adding the frequency of occurrence of each subject heading which, in alphabetical arrangement, had an occurrence of two or more. The corresponding 'most frequent class notation' is the frequency of the most identical class notation corresponding to the frequency of any given subject heading.

This process was repeated for all subject headings in four categories of bibliographic records, for all subject headings with different orders in bibliographic records, and with respect to both DDC and LCC notations

(a) LCSH Versus LCC Notations

Using 'List #1', the frequency of each subject heading which had an occurrence of two or more was counted. Then the frequency of the most frequent identical class notation corresponding to each subject heading was determined. This analysis was done with respect to the number of subject headings per bibliographic record. To find probability of each category, the sum of frequencies for LCC notations was divided by the sum of LCSH frequencies of the same category. In addition, the correlation measures were computed for each category of bibliographic records. The results of this analysis are summarized in Table 3.

TABLE 3
PROBABILITY AND CORRELATION BETWEEN SUBJECT HEADINGS AND LCC CLASS NOTATIONS IN FOUR GROUPS OF BIBLIOGRAPHIC RECORDS.

Groups of Bibliographic Records	Subject Frequency	LCC Frequency	p	r	Significance of r
1SH	133	108	.81	.93	.001
2SH	206	106	.50	.66	.001
3SH	51	20	.39	.68	.001
4SH or more	20	12	.60	.77	.001
Total*	776	347	.45	.46	.001

* In this table and the following tables whenever the 'total' is designated by an astrisk it indicates that the figures of the total are not the sum of previous categories, they are the actual frequency count of the variables involved.

There was a highly strong correlation (r=0.93) between subjects and corresponding LCC class notations in the first category of bibliographic records. A t test for correlation was significant at the 0.001 level for all categories. It should be noted that in spite of the fact that probability of the third group is less than the second, the correlation measures for the third is higher than the second. Both probability and correlation measures for the fourth group are higher than those of the third.

The difference between the four groups of bibliographic records was tested by chi square. The results of this test are shown in Table 4. The test confirmed that the difference between

four groups of bibliographic records is significant at the 0.01 level. The calculated chi square (16.143) was higher than the table value (11.345, df=3) at the 0.01 level.[19] Therefore, the null hypthesis of no difference was rejected at alpha level 0.01. The difference between the third and fourth groups was not significant. The calculated chi square (0.958, without Yates's correction for continuity this difference was 1.381) was less than table value (1.074, df=1) at the 0.30 level[20]. That is, the higher probability for the fourth category of bibliographic records could be by chance alone. With a relatively small size of the sample, and the fact that the sample was not totally random, this difference may be due to sampling error.

TABLE 4
CHI SQUARE TEST OF DIFFERENCE BETWEEN FREQUENCIES OF LCC NOTATIONS IN FOUR GROUPS OF BIBLIOGRAPHIC RECORDS.

Groups	Frequencies *			Chi Square Values **			
	LCSH	Observed LCC	Expected LCC	Test Value	df	Table Value	s
1 SH	133	108	79.80				
2 SH	206	106	123.60				
3 SH	51	20	30.60				
4 SH	20	12	12.00				
Total	410	246	246	16.216	3	11.345	.01

* In this test and the following chi square test the expeted frequency in each cell was calculated with respect to the value of its corresponding independent variable in total observations (e.g. 246 x 133 / 410 = 79.80).
** df = degrees of freedom s = significance

The results of the above analysis indicated that the degree of association between LCSH and LCC notations decreases as the number of subject headings per document increases. The degree of variability of LCC with respect to the number of subjects per bibliographic record is given by the following regression equation for LCSH and LCC. The regression equation indicates that changes in LCC frequency largely depends on the changes in the first group of bibliographic records:

$$LF = 0.48 + 0.75(SH1) + 0.31(SH2) + 0.34(SH3) + 0.43(SH4)$$

Where LF is the frequency of Library of Congress Classification and SH1, SH2, SH3, and SH4 stand for frequency of subject headings in the four groups of bibliographic records.

(b) LCSH Versus DDC Notations

Similarly, 'List #2' was used to determine the frequency of any LCSH, which had an occurrence of two or more, and the frequency for the most frequent DDC notation corresponding to each subject heading. Then, the probabiliy and correlation of LCSH with respect to DDC notations were calculated. Table 5 represents the results of probability and correlation measures between subject headings and DDC class notations in four groups of bibliographic records.

TABLE 5
PROBABILITY AND CORRELATION BETWEEN SUBJECT HEADINGS AND DDC NOTATIONS IN FOUR GROUPS OF BIBLIOGRAPHIC RECORDS.

Bibliographic Records	Subject Frequency	DDC Frequency	p	r	Significance of r
1SH	133	88	.66	.85	.001
2SH	214	91	.43	.59	.001
3SH	51	18	.39	.65	.001
4SH or more	20	12	.60	.77	.001
Total*	776	298	.38	.38	.001

A t test for correlation indicated that all the correlations were significant at alpha level 0.001. There was a correlation of 0.85 between subject headings and their corresponding DDC class notations in the group of bibliographic records with one subject heading, which is a strong correlation.

Similar procedures reported in Table 4 were followed to test the difference between DDC frequencies in four groups of bibliographic records. For sake of brevity the tables constructed for testing the differences are not repeated in the published report, rather for each test only the results are reported. The chi square test confirmed a significant difference between the four groups of bibliographic records. The caculated chi square (11.950) was higher than table value (11.345, df=3) at the 0.01 level. However, as in the case of LCC, the difference between the fourth group (i.e. 4SH) and the third (1.533) was not significant at the 0.20 level (table value = 1.642, df=1). Therefore, the fact that the probability for fourth group is higher could be by chance alone.

The results of this analysis confirmed that: as the number of subject headings per bibliographic record increases, the degree of association between LCSH and DDC notations decreases. A regression analysis of LCSH and DDC with respect to the number of subject per bibliographic record indicated that the variability in the first group of the bibliographic records has

the highest share in the variability of the DDC frequency. The variablity of DDC frequency with respect to the number of subject headings can be stated in the following equation:

$$DF = 0.40 + 0.53(SH1) + 0.35(SH2) + 0.33(SH3) + 0.47(SH4)$$

where DF is the frequency of Dewey Decimal Classification notations and SH1, SH2, SH3 and SH4 represent the groups of bibliographic records with one Subject heading, two subject headings, etc..

2. Changes in Order of Subject Headings Versus Class Notations

The second question was to see if the degree of association between subject headings and their corresponding class notations varies with respect to the order of subject headings in the bibliographic records. The probability and Pearson's correlation measures were calculated with respect to the order of subject headings in the four groups of bibliographic records to find out the answer to this question.

(a) LCSH Versus LCC Notations

Using 'List #1', the degree of association between LCSH and LCC was analyzed with respect to the changes in the order of subject headings. The results of the analysis showed that

whenever the order of subject headings increases, both probability and correlation measures decrease. The changes in the probability and correlation between subject headings and LCC notations with regard to the changes in the order of subject headings are given in Table 6.

TABLE 6
CHANGES IN PROBABILITY AND CORRELATION BETWEEN SUBJECT HEADINGS AND LCC NOTATIONS WHEN THE ORDER OF SUBJECT HEADINGS CHANGES.

Subject Order	Subject Frequency	LCC Frequency	p	r	Significance of r
1	367	235	.64	.80	.001
2	114	35	.31	.52	.001
3	17	4	.24	.46	.001
4	0	0	-	-	-
Total*	776	347	.45	.46	.001

The difference between orders 1, 2, and 3 was tested by chi square (since frequencies for the fourth order were zeros it was not included in chi square test). The test showed that there is a significant difference between LCC frequencies corresponding to subject headings with different orders. The calculated value of chi square was 20.732 as compared with table value of 13.815 for df=2 at the 0.001 level.

The regression analysis of LCSH and LCC with respect to the order of subjects in the bibliographic records indicates that as the order of subject headings increases, the changes in subject headings have less effect on the variability of LCC frequencies. The regression equation of LCC and LCSH with respect to the changes in order is given in the following equation:

$$LF = 0.37 + 0.57(S1) + 0.16(S2) + 0.14(S3)$$

Where LF is the frequency of the Library of Congress Classification, and S1, S2, S3 are groups of subject headings with order of one, two, and three. This equation indicates that subject headings with order of one have the highest share of variability in LCC frequencies.

The changes in probability and correlation measures of LCSH and LCC, when both number of subject headings per bibliographic record and the order of subject headings change, are given in Table 7. This table shows that whenevere the number of subject headings per bibliographic record and the order of subject headings increase both both probability and correlation measures decrease.

TABLE 7
VARIATION IN DEGREE OF ASSOCIATION BETWEEN SUBJECT HEADINGS AND LCC NOTATIONS WITH RESPECT TO THE CHANGES IN ORDER AND NUMBER OF SUBJECT HEADINGS PER BIBLIOGRAPHIC RECORD.

SH	1SH		2SH		3SH		4SH	
ORDER	p	r	p	r	p	r	p	r
1	.81	.93	.66	.83	.50	.81	-	-
2	-	-	.34	.57	.50	.57	-	-
3	-	-	-	-	.14	.37	-	-
4	-	-	-	-	-	-	-	-

Again, using the actual frequencies of LCC notations, a chi square test indicated that there is no significant difference in the first row. For the first row, i.e. subject headings from three groups of bibliographic records with order of one, the calculated chi square (2.776) was less than table value (3.219, df=2) at the 0.20 level and the null hypothesis of no difference was accepted. This test indicates that regardless of the number of subject headings per bibliographic record, when the order of subjects are the same the degree of association between subject headings and their corresponding class notations will be the same. However, due to insufficient data and the fact that expected value for some cells were less than five the test of difference for the subsequent rows was not possible.

In group of the bibliographic records with two subject

headings, there was a significant difference between the first subject headings and the second subject headings. A Chi square of 5.732 was calculated which was greater than the table value (5.412, df=1) at the 0.02 level. Due to insufficient data in some of the cells in 3SH, the test of difference between three orders in this groups was not possible.

(b) LCSH Versus DDC Notations

'List #2' was used for the analysis of the changes in the order of subject headings and any possible variations of association between LCSH and DDC class notations. The results of order analysis showed that both probability and correlation measures between LCSH and DDC notations decrease as the order of subject headings increases. The variations of association between subject headings and class notations with respect to the order of subject headings are given in Table 8.

TABLE 8
VARIATION IN PROBABILITY AND CORRELATION OF SUBJECT HEADINGS AND DDC NOTATIONS WHEN ORDER OF SUBJECT HEADINGS CHANGES.

Subject Order	Subject Frequency	DDC Frequency	p	r	Significance of r
1	367	193	.53	.70	.001
2	114	29	.25	.46	.001
3	17	6	.35	.56	.001
4	0	0	-	-	-
Total*	776	298	.38	.38	.001

A substantial correlation (r=0.70) between subjects and DDC notations was found for the category of subject headings with order of one. A chi square test indicated a significant difference in the frequencies of DDC with respect to the order of subject headings. The calculated chi square was 14.425 which was higher than the table value (13.815, df=2) at the 0.001 level. However, both probability and correlation measures for the subject headings with order of three, were more than the probability and corrolation measures for subject headings with order of two. It may be due to the chance , but due to insufficient data it was not possible to test the significance of the differences.

The regression analysis of LCSH and DDC notations with

respect to the changes in the order of subjects showed that the subject headings with order of one have the highest share in variability of DDC as shown in the following equation:

$$DF = 0.40 + 0.41(S1) + 0.11(S2) + 0.27(S3)$$

where DF is the frequency of Dewey Decimal Classification notations and S1, S2, and S3 are groups of subject headings with order of one, two and three.

The changes in probability and correlation measures of LCSH and DDC, when both order and the number of subject headings per bibliographic record change, are given in Tables 9. Both of these measures decrease as the number of subject headings per bibliographic record increases and/or order of subject headings increases.

TABLE 9
VARIATION IN DEGREE OF ASSOCIATION BETWEEN SUBJECT HEADINGS AND DDC CLASS NOTATIONS WHEN BOTH ORDER AND THE NUMBER OF SUBJECT HEADINGS PER BIBLIOGRAPHIC RECORD CHANGE.

SH	1SH		2SH		3SH		4SH	
Order	p	r	p	r	p	r	p	r
1	.66	.85	.48	.68	.38	.65	-	-
2	-	-	.24	.49	.50	.57	-	-
3	-	-	-	-	.29	.53	-	-
4	-	-	-	-	-	-	-	-

Using the actual frequency figures of DDC notations in the above table, a chi square test showed that there is not a statistically significant difference between frequencies for the first row; the calculated chi square (3.638) was not significant at the 0.10 level (table value=4.605, df=2). This test indicated that as long as the order of subject headings is one, it does not matter whether it is coming from a bibliographic record with one, two,or three subject headings. Due to insufficient data, the difference test for order 2 and three was not possible.

The difference between orders one and two in 2SH category was significant at the 0.05 level. The calculated difference was 4.422, which was greater than the table value (3.841, df=1). However, for 3SH the test of difference was not possible because the data were not sufficient.

3. Comparison of LCC and DDC

The third question was to see if the LCC and DDC classification systems are significantly different with respect to the degree of association between LCSH and their corresponding class notations in the two systems. Table 10 represents the comparison of the two classification with respect to the number of subjects per bibliographic records. In general, LCC frequencies were higher than the DDC frequencies in the four

groups of bibliographic records. However, in spite of same frequency in groups three and four in both classifications, the correlation measure for LCC in group 3SH is higher than the correlation of DDC for the same group.

The results of a chi square test showed that the differences between scores of LCC and DDC with respect to the association with their corresponding subject headings in four categories of bibliographic records are not statistically significant at the 0.05 level.

TABLE 10
COMPARISON OF PROBABILITY AND CORRELATION MEASURES BETWEEN LSCH AND THE TWO CLASSIFICATION SYSTEMS WHEN THE NUMBER OF SUBJECT HEADINGS PER BIBLIOGRAPHIC RECORD CHANGES.

Groups	Frequencies			p	p	r	r	Chi Square Values			
	SH	DDC	LCC	DDC	LCC	DDC	LCC	Test	df	Table	s
1SH	133	88	108	.66	.81	.85	.93	1.842	1	1.642	.20
2SH	214	91	106	.43	.50	.59	.66	.995	1	.455	.50
3SH	51	20	20	.39	.39	.65	.68	No Difference			
4SH	20	12	12	.60	.60	.77	.79	No Difference			
Total*	776	298	347	.38	.45	.38	.46	3.572	1	2.706	.10

The comparison of the two classification systems with respect to the changes in the order of subjects is given in Table 11. This table indicates that in all cases the values for LCC

are higher than those for DDC. However, a chi square test showed that only the difference between scores of two classification systems in the group of subject headings with order one was statistically significant at the level 0.05. As shown in this table, there were no significant difference between values in order two, and between values in the total.

TABLE 11
COMPARISON OF PROBABILITY AND CORRELATION MEASURES BETWEEN LCSH AND THE TWO CLASSIFICATION SYSTEMS WHEN THE ORDER OF SUBJECT HEADINGS CHANGES.

Order	Frequencies			p	p	r	r	Chi Square Values			
	SH	DDC	LCC	DDC	LCC	DDC	LCC	Test	df	Table	s
1	367	193	235	.53	.64	.71	.80	3.928	1	3.841	.05
2	114	29	35	.25	.31	.46	.52	.391	1	.102	.75
3	17	6	4	.35	.24	.56	.46	.100	1	.062	.80
4	0	0	0	-	-	-	-	No Data			
Total*	776	298	347	.38	.45	.38	.46	3.572	1	2.706	.10

SUMMARY AND CONCLUSION

1N this study 1,893 Library of Congress MARC bibliographic records, which had a Dewey Decimal Classification of 500-699, were analyzed to measure the degree of association between the Library of Congress subject headings and their corresponding class notations. These bibliographic records were differentiated by the number of subject headings which have been assigned to them by the Library of Congress Cataloging Division, and the order of subject headings in the bibliographic records. There were 3,280 subject headings in the retrieved bibliographic records. Using different sort key order, these subject headings and their corresponding class notations were sorted in two different lists. Each list was analyzed to find the answers to the research questions in this study.

Based on the analysis of the two lists and calculation of probability, correlation measures, regression equations, and chi square tests of differences, the following conclusions can be drawn from this study:

1. With reference to the first question, whenever the number of subject headings per bibliographic record increases, in both LCC and DDC systems, the degree of association between

subject headings and their corresponding class notations decreases. In terms of the association between subject headings and their corresponding class notations, there were significant differences between four groups of bibliographic records. This was true for both DDC and LCC.

2. With regard to the second question, it was found that in both classification systems, whenever the order of subject headings in bibliographic records increases, the degree of association between subject headings and their corresponding class notations decreases. The subject headings with order of one had a higher probability and correlation than those subject headings with order of two or more.

3. With respect to the third question, generally the correlation and probability measures between LCSH and LCC were higher than the same measures for LCSH versus DDC. However, chi square tests indicated that, except for subject headings with order of one, these differences were not significant. The two classification systems were not significantly different with respect to the measures of association in this study at the 0.05 level.

The sample size of this study does not permit to make a generalization. However, based on the above tests it is recommended that in collections which have science and technology

monographs, it would be more appropriate to have some kind of hierarchy in filing subject headings in such a way that subject headings with order of one, in the bibliographic record, be filed before the same subject headings with order of two or more. It is true that when a given subject heading is assigned to a bibliographic record, the contents of the document being described has elements of that subject; but, as this study suggests, those bibliographic records which have one subject heading have a higher correlation with their corresponding class notation. Similarly, in bibliographic records with multiple subject headings, those subject headings being listed first have higher correlation with their corresponding class notations than headings that happens to be second, or third subject headings.

It is recommended that by using bibliographic records in all fields of knowledge, a more comprehensive study be conducted. This study was limited to science and technology monographs and this limitation may have some effects on the findings. The size of sample in this study was not large enough to have sufficient occurrences of subject headings in all categories of bibliographic records, this may have some effects on the outcome of the study. The data was relatively old, and it can be regarded as a limitation, though it may be argued that since both subject headings and class notations change over a period of time, it does not affect their degree of association in any given time. However, this is a matter of further investigation.

Any future study should have a sample size large enough to allow sufficient frequencies in each category. Having sufficient occurences of subject headings in categories that in this study had fewer frequencies, such as category of bibliographic records with four subject headings, would allow a more justified conclusion for the statistical analyses. There may also be a comparison between degree of association between various main classes in both classification systems to see if the degree of association between variables differs from one main class to another.

REFERENCES

1. Bohdan S. Wynar. Intoduction to Cataloging and Classification, 5th ed. Prepared with assistance of John Phillip Immroth (Littleton, Colo., Libraries Unlimited, 1976) Pp. 329-330.

2. Jay E. Daily. "Subject Headings and the Theory of Classification," American Documentation 8 (October 1957):269-274.

3. Ibid., p. 272.

4. Martha L. Manheimer. "The Relationship of Classified Library of Congress Subject Headings to the Library of Congress Classification Schemes," In (Williams, James G. , Martha L. Manheimer, and Jay E. Daily eds. Classified Library of Congress Subject Headings, New York: Marcel Dekker, 1972, vol. 1) Pp. 13-25.

5. Ibid., p. 15.

6. R. A. Fairthorne. Towards Information Retrieval; Index compiled by Calvin N. Mooers (London: Butterworths, 1961) p. 95.

7. John S. Davis and M. F. Haymes. "Organization of a Geographic Data Collection," Special Libraries 69 (May-june 1978): 215-219.

8. "Order of Subject Tracings," Cataloging Service Bulletin Number 1 (Summer 1978):15.

9. Melvil Dewey. Dewey Decimal Classification and Relative Index; Ed. 19, edited under the direction of Benjamin A. Custer (Albany, Forest Press, 1979, vol.2) p. 1179.

10. Library of Congress Subject Cataloging Division. Classification. Class T: Technolgoy, 5th ed. (Washington, The Library of Congress, 1971) P. 130.

11. Melvil Dewey, p. 1180.

12. Phyllis Allen Richmond. "Cats: An Example of Concealed Classification in Subject Headings," Library Resources and Technical Services 3 (Spring 1959):102-112.

13. Harry Dewey. "The Relationship Between the Headings in the Subject Catalog and the Classification Numbers of the

Books," In (Perreault, Jean M. ed. Reclassification: Rationale and Problems; Proceedings of a Conference on Reclassification, held at the Center of Audult Education , University of Maryland, College Park, April 4 to 6, 1968. College Park, School of Library and Information Services, University of Maryland, 1968) Pp. 57-78.

14. Manheimer, p. 16.

15. Ibid.,Pp. 19-20.

16. John Philip Immroth, Analysis of Vocabulary Control in Library of Congress Classification and Subject Headings (Littleton, Colorado: Libraries Unlimited, 1971) Pp. 144-145.

17. Edward O'neill and Rao Aluri. Research Report on Subject Heading Patterns in OCLC Monographic Records, (Colombus,Ohio, OCLC,Inc., Research and Development Division, 1979). See also the same report under the tile of "Library of Congress Subject Heading Patterns in OCLC monographic Records" in Library Resources & Technical Services 25 (January/March 1981):63-80.

18. Jan Wepsiec. "Languge of the Library of Congress Subject Headings Pertaining to Society," Library Resources and Technical Services 25 (April/June 1979): 197-203.

19. Chi square table values reported in this paper are from Dennis J. Palumbo Statistics in Political and Behavioral Science; Revesided Edition (New York: Columbia University Press, 1977). Table 3.

20. Hereafter, for any test with one degree of freedom, only the corrected value is reported.